

1. Pomiary gęstości Ziemi (w g/cm^3) (obecnie przyjmuje się wartość $5,515 \text{ g}/\text{cm}^3$), wykonane przez Cavendisha w 1798 za pomocą wagi skręceń:

4,07	4,88	5,10	5,26	5,27	5,29	5,29	5,30	5,34	5,34
5,36	5,39	5,42	5,44	5,46	5,47	5,50	5,53	5,55	5,57
5,58	5,61	5,62	5,63	5,65	5,75	5,79	5,85	5,86	

- Dlaczego pomiary *jednej* stałej fizycznej dają tak różne wartości?
 - Zaproponuj wartość gęstości Ziemi na podstawie pomiarów Cavendisha. Czy zaproponowana przez ciebie wartość jest różna od $5,515$? Dlaczego?
 - Sformułuj warunki, jakie według ciebie, powinna spełnić wartość, reprezentująca zbiór pomiarów.
2. Liczba wypadków samochodowych, gdy kierowcy używali telefonu komórkowego:

pon	wto	śro	czw	pią	sob	nie
133	126	159	136	113	12	20

Zaproponuj wartość charakteryzującą liczbę wypadków, przypadającą na jeden dzień tygodnia.

Wartość typowa danych.

Wartość typowa danych jest możliwie najbliższa wszystkim dostępnym danym.

Niech x będzie n -wymiarowym wektorem danych, d - odległością w przestrzeni \mathcal{R}^n . **Wartością typową danych** x względem metryki d jest liczba:

$$m(x, d) = \arg \min \{d(x, u1_n) : u \in \mathcal{R}^n\},$$

gdzie 1_n jest wektorem, składającym się z n jedynek.

3. Wyznacz wartość typową danych x dla odległości euklidesowej $d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$.

Popularne odległości w \mathcal{R}^n

Odległość Hamminga

$$d_H = \sum_{i=1}^n 1[x_i \neq y_i],$$

gdzie $1[A] = 1$ gdy warunek A zachodzi i $1[A] = 0$ gdy warunek A nie zachodzi.

Odległość Czebyszewa

$$d_C(x, y) = \max \{|x_i - y_i| : i = 1, 2, \dots, n\}$$

Odległość taksówkowa

$$d_1(x, y) = \sum_{i=1}^n |x_i - y_i|$$

4. Wyznacz wartość typową danych x dla odległości Hamminga, Czebyszewa i taksówkowej.

5. Oblicz wartości typowe względem odległości euklidesowej, Hamminga, Czebyszewa i taksówkowej dla danych z zad. 1-2. Skomentuj otrzymane wyniki. Które z wartości typowych najlepiej oddają charakter danych z zad.1, a które danych z zad.2?

Rozrzut danych

Rozrzut danych charakteryzuje koncentrację danych wokół wartości typowej.

$$s(x, d) = d(x, m(x, d) 1_n)$$

6. Oblicz $s(a1_n, d)$

7. Wyznacz rozrzut względem odległości euklidesowej, Hamminga, Czebyszewa i taksówkowej¹.

8. Oblicz rozrzut względem odległości euklidesowej, Hamminga, Czebyszewa i taksówkowej dla danych z zad. 1-2.

¹ Wielkość $1/\sqrt{n}s(x, d_2)$ nazywa się odchyleniem standardowym.